The datafication of teaching in Higher Education

The datafication of teaching in Higher Education: critical issues and perspectives

Ben Williamson, Sian Bayne and Suellen Shay

Introduction

Contemporary culture is increasingly defined by data, indicators and metrics. Measures of quantitative assessment, evaluation, performance, and comparison infuse public services, commercial companies, social media, sport, entertainment, and even human bodies as people increasingly quantify themselves with wearable biometric devices. In a 'society of rankings', simplified and standardized metrics act as key reference points for making sense of the world (Esposito & Stark, 2019, 15). Beyond conventional statistical practices, the availability of 'big data' for large-scale analysis, the rise of data science as a discipline and profession, and the development of advanced technologies and practices such as machine learning, neural networks, deep learning and artificial intelligence (AI), have established new modes of quantitative knowledge production and decision-making (Kitchin, 2014; Ruppert, 2018).

Although 'datafication'—the rendering of social and natural worlds in machine-readable digital format—has most clearly manifested in the commercial domain, such as in online commerce (e.g. Amazon), social media (Facebook, Twitter), and online advertising (Google), it has quickly spread outwards to encompass a much wider range of services and sectors. These include, controversially, the use of facial recognition and predictive analytics in policing, algorithmic forms of welfare allocation, automated medical diagnosis, and—the subject of this special issue—the datafication of education.

Education is a particularly important site for the study of data and its consequences. The scale and diversity of education systems and practices means that datafication in education takes many forms, and has potential to exert significant effects on the lives of millions. That education is widely understood as a public good, rather than a commercial enterprise (with some exceptions) also means that the extraction of data from students, teachers, schools and universities cannot be straightforwardly analysed as another instantiation of 'surveillance capitalism', that is, the gathering of the 'raw material' of human life *en masse* for analysis, sale and profit (Zuboff, 2019). Instead, the datafication of education needs to be understood and analysed for its distinctive forms, practices and

consequences. In this brief editorial introduction to the special issue on 'The datafication of teaching in higher education', we situate the papers in wider debates and scholarship, and outline some key cross-cutting themes.

Measurement matters

There is of course a very long history to practices, processes and technologies of datafication in which current developments in big data, AI and machine learning need to be situated (Beer, 2016). The eighteenth and nineteenth centuries witnessed an outpouring of statistical knowledge production, as everything from industrial manufacturing to the natural world, and from the state of the human population to the workings of the human body itself, was subjected to quantification and increasing numerical management (Bowker, 2008; Ambrose, 2015). The work of modern government itself came to rely on statistics, as people, goods, territories, processes and problems were all made legible as numbers, and statistical knowledge came to 'describe the reality of the state itself' (Foucault, 2007, 274) as part of the 'machinery of government' (Rose, 1999, 213).

The statistical machinery of the nineteenth- and twentieth-century state is now, in the twenty-first century, shadowed by a vast complex of data infrastructures, platforms, devices, and analytics organizations from across the public, charitable and private sectors, as big data has itself become a new source of knowledge, governance and control (Bigo, Isin & Ruppert, 2019). Social media platforms, web interactions, financial transactions, public surveillance networks, online commerce, business software, mobile phone location services, wearable devices, and even connected objects in the Internet of Things have become key sources of knowledge for those authorities with access to the data they produce (Marres, 2017). Governments are increasingly turning to digital services in order to generate detailed information about the populations they govern, including controversial attempts to introduce public facial recognition systems for purposes of individual identification (Crawford & Paglen, 2019). Through machine learning, neural nets and deep learning, so-called AI products and platforms can now 'learn from experience' in order to optimize their own functioning and adapt to their own use (Mackenzie, 2018). Nineteenth- and twentieth-century 'trust in numbers' has metamorphosed into a 'dataist' trust in the 'magic' of digital quantification, algorithmic calculation, and machine learning (Elish & boyd, 2018)

Dataism is a style of thinking that is integrally connected to processes of neoliberalization, as competitive logics and the desire to compare the performance of entities against each other, as if they are competing in markets, have been incorporated into various forms and technologies of measurement. Beer (2016, 31) argues that this period of intensive quantification is governed under a particular neoliberal system of 'metric power', and that 'understanding the intensification of

measurement, the circulation of those measures and then how those circulations define what is seen to be possible, represents the most pressing challenge facing social theory and social research today'. He suggests a number of key themes for understanding metric power (Beer, 2016, 173-77). Data and metrics set *limits* on what can be known and what can be knowable. They define what is rendered *visible* or left invisible, thereby impacting on how certain practices, objects, behaviours and so on gain value, while others are not measured or valued. Measurement involves *classification,* sorting, ordering, and categorizing people and things, which defines how they are known and treated. It leads to *prefiguring judgment*, by setting desired aims and outcomes with the aim to bring the future into the present, which a measurement is designed to help achieve. Data-based processes also e*xpand* into new tasks, functions and programs, and *intensify* their influence. The intensification of measurement leads to forms of *authorization* and endorsement of certain outcomes, people, actions, systems, and practices, thus marking out what is claimed to be truthful. It also involves increasing *automation,* which shapes human agency and decision-making—automated systems of computation are taken as objective, legitimate, fair, neutral and impartial, and impact on human judgement. Finally, metrics induce *affective* reactions, such as anxiety or competitive motivation, and thereby promote or produce actions, behaviours, and pre-emptive responses by prompting people to perform in ways that can be valued, compared and judged in measurable terms.

The power of metrics to affect how social and natural worlds are known and compared, and therefore to shape how they are treated and changed, means that measurement matters. Data and metrics do not just reflect what they are designed to measure, but actively loop back into action that can change the very thing that was measured in the first place. Data practices materialize the competitive neoliberal impulse to ensure efficient market functioning and constant improvement through measurement, the hierarchization of winners and losers, and the attribution of quantitative value. This can be fairly mundane, in the case of an online retailer recommending future goods to purchase based on past purchasing record and comparison against millions of other shoppers, where the measured market is the source of the recommendation. Media streaming services constantly capture data about consumption habits, and feed that back into recommended shows and playlists. The metrics in such cases include favoured genres, time spent listening or watching, artists or shows selected and so on. This may seem fairly banal, but yet is shaping cultural habits and individual tastes. But measurement also matters for even more consequential reasons. It has changed the ways economies function, serving hypercapitalist objectives of making data into a key source of market value (Fourcade & Healy, 2017). Surveillance systems such as predictive policing and facial recognition disproportionately focus suspicion on ethnic

minority groups, and reinforce longstanding structural inequalities in societies (Crawford & Paglen, 2019), as judgments are made based on various forms of comparison and prediction.

Education has long been subject to historical forms of 'datafication' (Lawn, 2013), but the quantification, measurement, comparison, and evaluation of the performance of institutions, staff, students, and the sector as a whole is intensifying and expanding rapidly. Higher Education is itself implicated in neoliberalizing forms of metric power, as various technologies of data-based measurement and evaluation impose limits on what is made visible and known, sort people and outcomes into (sometimes hierarchical) categories, establish measurable aims, expand to new tasks, establish what is claimed to be true or valuable, impose automation on decision-making, and affect the ways people feel, act and behave.

The datafication of education

As part of the expansion of metric power across social, cultural, economic and political processes, the education sector as a whole has experienced a dramatic increase in datafication. Demands of measurable accountability, international large-scale assessments, comparative performance benchmarking, and the proliferation of metrics and indicators from the early years through schooling and higher education, are all part of a long genealogy of 'governance by numbers' (Piattoeva & Boden, 2020). Amid wider social, economic and political enthusiasm for 'big data' and 'AI', technologies such as learning analytics, adaptive 'personalized learning' platforms, and robot teaching assistants have all been developed, promoted and taken up by education institutions (Williamson, 2017). Datafication in education has even begun to expand to systems such as facial recognition (Andrejevic & Selwyn, 2019) and 'emotion AI' based on wearable biosensors, body gesture and facial expression analysis (McStay, 2018).

More specific to the higher education sector, the collection and use of university data has expanded and mutated dramatically in scope over the last two decades, as the result of significant efforts by political centres and supporting businesses, think tanks, consultancies and sector agencies (Williamson, 2018). Research metrics are used to audit, compare, and assess the quality of research outputs and impact (Wilsdon et al, 2015), and the rating of university teaching quality and the 'value' of academic labour has increased 'quantified control' and 'metricization of the academy' (Burrows, 2012, p. 356). University rankings and league tables produce new kinds of reactive behaviours, as institutions and individuals seek out ways of maximizing their performance in terms of the measures they are scored on (Espeland & Sauder, 2017). Digital technologies and interests in 'big data' have now enlarged the scope of measurement across education systems, increased the

fidelity of data analyses, and enhanced the uptake and use of data for various forms of audit, inspection, evaluation and decision-making.

A political shift towards the collection of student data is catalysing significant technological development and organizational activity. It is focusing the sector far more closely on measuring and comparing students' 'learning gain', their 'engagement' in their studies, their 'satisfaction' with the 'student experience', and on the overall 'quality' of learning provision in different institutions and degree programmes. Technologies such as predictive learning analytics and 'AIed' applications, originally produced in university research centres and labs, have crept across the sector to perform these functions (Buckingham Shum & Luckin, 2019). Measures of student performance, sentiment, engagement, and satisfaction are also treated as proxy measures of the performance of staff, courses, schools, and institutions as a whole, leading to new claims that HE quality can be adduced from the analysis of large-scale student data (Williamson, 2019).

The datafication of HE through the evaluation machinery of research assessment exercises, teaching evaluation frameworks, impact measurements, student satisfaction ratings, and so on, is frequently viewed as part of an ongoing process of marketization of the sector (Busch, 2016). One particularly polemical critique describes a 'pathological organizational dysfunction' whereby corporate models of marketization, competition, audit culture, and datafication have combined to produce 'the toxic university' (Smyth, 2017). HE, it is argued, has been made to resemble a market in which institutions, staff and students are all positioned competitively, with measurement techniques required to assess, compare and rank their various performances. Educational technology advocacy has thrived in this context, with even the European Commission announcing in 2020—as part of its White Paper on Artificial Intelligence—plans to 'improve education' across Europe through the use of AI-based predictive learning analytics.

In parallel with political desires to subject HE to further datafication, a 'global HE industry' has emerged of 'data solutions' service providers and platform companies offering HE products, which have sought to open up and exploit new markets in HE data (Komljenovic & Robertson, 2018). Market projections value HE technologies as worth billions, and stimulate investment, venture capital, private equity and philanthropic interest. Education businesses have begun shifting to 'digital first' products as ways of creating new markets for their products, thereby making it possible for them to generate large-scale educational data for analysis, further product refinement and development (Sellar & Hogan, 2019). The parent companies of major global edtech products have been sold in complex multibillion dollar deals as edtech has become a major site of commercial advantage. And beyond these edu-businesses, other technology companies have expanded into HE

markets, including the global web giant Amazon, which has begun pushing its voice interface Alexa into universities as a communication intermediary between students and institutions. For businesses in the global HE industry, data are a key source of market-making, both as universities seek out new technologies to help them measure their performances, and the companies utilize those data for product refinement and further development (Williamson, 2020).

These commercial developments raise concerns over the monetization of student data, and particularly of how businesses have gained advantage from the use of that data for 'training' machine learning systems (Perrotta & Selwyn, 2019). New organizations have even suggested that it may be possible to quantify the value of every university module, course or career choice and, by consolidating a permanent record of students' qualifications and skills from across the whole educational 'supply chain'—as 'learner wallets' hosted on blockchain technologies—offer AI-enhanced employability advice and enable students to securely share their data with employers (Gent, 2020). Such blockchain-based programs are even supported by the US government, with educational records and outcomes data positioned as a form of currency that students can exchange for employment advantage and economic opportunity (ACE, 2020). The massive collection of student data, then, is seen as an opportunity not just to measure HE quality, provide insights for institutional improvement, or to inform edu-businesses' product development, but to lubricate the pipeline from 'learning to earning' and from 'major to wages'. The economic perspective on student data is buttressed and reinforced by initiatives to measure the 'value' of HE, such as the Postsecondary Value Commission established in the US by the Bill and Melinda Gates Foundation (www.postsecondaryvalue.org), and by 'graduate tracer studies' that track individuals from enrolment to their graduate destinations and earnings. In the UK, the Longitudinal Educational Outcomes dataset, which shows the median earnings of graduates from specific degree courses, is already used to help prospective students choose where and what to study, via the national Discover Uni website, and is also currently being considered as a way of rating universities as data become central to reshaping HE around 'employability' (Fazackerley, 2020).

Through these government priorities, platform companies, edu-businesses and initiatives, diverse forms of student data are being generated, analysed and used. They include everything from historical and socio-demographic records, library metrics, learning management system logins, assessment systems, and plagiarism detection software, to learning analytics traces, learning gain data, longitudinal graduate outcomes data, attendance and location records, and social media activity. Many sources of student data are being made interoperable, allowing extensive data-linking, time-series analysis, and aggregation of datasets, as they are connected into large-scale networked infrastructures for measuring, evaluating and governing

the performance of students, staff, schools, and whole institutions (Williamson, 2019).

The combined forces of neoliberal metric power, political reform and the global education industry are changing the ways higher education operates. University functions and processes are increasingly disaggregated or 'unbundled' into discrete services and tasks, often outsourced or fulfilled by third-party providers (McCowan, 2017), and then 'rebundled' into new components and models, many of which may be monetized by commercial companies in the competitive HE market (Czierniewicz, 2018). Bacevic (2018) argues that universities are 'assemblages' of different functions and objects that constantly shift and mutate in relation to changing social, political, economic and technological circumstances and forces. Contemporary efforts to embed data analytics and AI in HE are simply the latest instantiation of such ongoing dynamic transformations. Berry (2018), for example, suggests that social, political and economic forces associated with software, data and algorithms—characteristic of a 'data-intensive society'—are shaping a 'data-intensive university' that adapts in relation to its data environment and a new computational cultural milieu. As such, the contemporary university is reassembling into a new set of forms and functions as it adapts to a plethora of social, political, economic, and technological forces.

Datafication of teaching in HE

This special issue focuses critical attention on the datafication of teaching in higher education. Our point of departure is the understanding that as big data, analytics, and so-called AI technologies have swept across HE institutions in recent years—as a kind of experimental exemplar of wider social, technical and political forces—they have become key parts of a set of ongoing transformations that remain as yet uneven, messy, and contingent. Some positive accounts view datafication as a positive step towards a new kind of 'smart university' where data and automated technologies become integral to the everyday functioning of institutions, and play key parts in curriculum organization, pedagogy, student feedback, and assessment (Lane, 2014). Others take a much more downbeat view that HE datafication and automation open up universities, staff and students to new forms of surveillance, enhanced marketization and commercialization, and intensified performance measurement, management and governance (Castaneda & Selwyn, 2018).

The year 2020 is a good time to take stock of the various social, political and technical developments that are enabling the datafication of teaching in HE. Without wishing to rehearse the arguments of the papers collected in this issue in detail—which we invite you to read—we want to round out this editorial with some key themes emerging from the collection, and that deserve continued investigation and reflection.

Data power

The emerging field of 'critical data studies' has established the important understanding that data are not neutral. Despite being presented as impartial, objective and innocent, all data have to be produced and therefore bear the imprint of their producers (Kitchin, 2014). Choices have to be made about what data to collect, and how to analyse it. Computer technologies need to be programmed, checked, tuned, and maintained. These practical actions are framed by wider factors such as business plans, political objectives, or research aims. Data do not simply represent the reality of the world independent from human thought but are constructions about the world that have been assembled for specific purposes (Jasanoff, 2018). Thus 'data politics' operates in two ways: through struggles over data production and its deployments, and through the generation of new power relations at various scales (Bigo, Isin & Ruppert, 2018).

Prinsloo (this issue) makes the important point that commercial education businesses and government education departments are increasingly penetrating into education systems in the Global South, as businesses and governments in the Global North and east Asia seek out new spaces of profit and control. Drawing on Beer (2019), he argues that a 'data frontier' is being pushed ever further into education to reveal new authoritative insights and construct 'truths' about education, teachers and students. As such, those organizations that possess the means of datafication have invested in themselves particular forms of data power—the capacity to categorize and define what 'counts' as quality education, a good student or an effective teacher. Such systems, for Prinsloo, amount to a new form of 'data colonialism', especially where they originate in educational thinking and product development in the Global North. Where historical colonialism was concerned with the appropriation of territory and resources, and with rule over colonized subjects, data colonialism involves the exploitation of human beings through data, where continuous tracking leads to unprecedented opportunities for social discrimination and behavioural influence (Couldry & Meijas, 2018).

One compelling example of data colonialism in action in education is the Global Partnership for Education (GPE), an international organization that established a 'Data Solutions' program to introduce new data systems, and develop data skills capacity, in developing countries lacking 'sufficient data on education indicators' required for international comparison and benchmarking (https://www.globalpartnership.org/what-we-do/data-systems). With support from north American technology businesses, GPE is seeking to cast a statistical grid over education systems in the developing world, and to use the data to measure, compare, and improve them to 'global standards' of quality education. It is pushing the data frontier ever-further into quantitatively unknown territories, and raising significant challenges for curriculum and pedagogy.

Pedagogic effects

What is the pedagogy of datafication? There are two ways of approaching this question. First, many modern data systems with so-called AI capacity are built upon techniques of machine learning, neural networks and deep learning. These systems need to be taught and trained to learn in order to provide automated analysis of data and feedback (Knox, Williamson & Bayne, 2020). There is a kind of pedagogy involved in preparing and socializing AI to learn human subjects effectively. But datafication also demands certain kinds of classroom pedagogies. Not all forms of learning can be quantified and analysed. And this means, potentially, that not all forms of teaching and learning will 'count' in terms of how teachers and students are measured and assessed. The personalized learning platform built by Facebook engineers for the Summit Schools charter chain in the US, for example, was recently rejected by many students and parents because, they argued, it required extensive use of computers and foreclosed opportunities for pedagogic dialogue or critical independent thinking. 'Robot teaching' reduces teachers' pedagogic discretion and professional judgment.

Datafication brings the risk of pedagogic reductionism as only that learning that can be datafied is considered valuable. Brown (this issue) highlights how technologies such as learning analytics dashboards impose particular limits on how educators 'see' students. The dashboard, in other words, introduces limitations on the visibility that educators can attain over the human subjects in their classrooms. This is not to suggest that dashboards decrease visibility—after all, a key argument is that in large (often online) programs, educators already have very limited view of what students are doing—but that they direct the educator's attention to particular features of student activity and learning. As such, we see attention drawn to features such as 'engagement' and 'risk of drop-out' which can be easily quantified and visualized through the activities of students on digital platforms. There is a clear risk here that pedagogy may be reshaped to ensure it 'fits' on the digital platforms that are required to generate the data demanded to assess students' ongoing learning. Moreover, as students are made visible and classified in terms of quantitative categories, it may change how teachers view them, and how students understand themselves as learners.

Data subjects

The datafication of human beings affects how they are understood, treated, and acted upon. The concept of the 'data double' usefully refers to how digital profiles can be created from the activities of individuals (Raley, 2013). These profiles, or shadows, then become the basis for various forms of analysis and calculation, which circle back into individual experiences. To use the social media streaming example, the data double captured inside the database is used to make recommendations, which affects the consumer experience outside the database

(Cheney-Lippold, 2011). The individual becomes a data subject, defined and characterized algorithmically by being sorted into categories and predicted outcomes.

The construction of data doubles in education is especially consequential since anything that is modelled inside the database then affects the potentially life-changing experience of teaching and learning. A prediction of future progress based on past outcomes could radically affect the future prospects of the student by foreclosing curriculum opportunities. Forms of algorithmic education, in other words, deeply affect data subjects. In their paper, Harrison, Davies, Bell, Goodley, Fox and Downing (this issue) draw attention to how datafication both affects teaching and learning and shapes subjectivities. They refer to a 'student data subjects' which are assembled from digital traces of educational activity. Teachers, too, are increasingly known, evaluated and judged through data, and come to know themselves as datafied teacher subjects.

This datafication of student and teacher subjects prefigures a potentially profound transformation in how students and teachers understand themselves and in how they are understood and managed as learners and professionals. As Marachi emphasises (this issue), where 'frictionless' data transitions are enabled between primary, secondary and tertiary education and even the employment contexts of individuals, the data subject risks becoming a lifelong 'shadow' with potential impact which may be far from benign. Marachi calls for greater awareness, routine interrogation of data-sharing practices and critical distance between higher education institutions and 'edtech' platform partners promising 'enhancement' through data processing, the constitution of data subjects and the promises of 'personalisation'. Such changes may also demand that educators and students develop critical skills of using and evaluating data.

Data literacies

The surge in data use in education has prompted widespread concern with the development of teachers' (and students') 'data literacy' (Mandinach & Gummer, 2016). This concern takes two main forms. On one hand, some claim teachers need the data skills to be able to use new data systems and to be able to comprehend and utilize the data they produce for purposes of professional improvement, assessment of students, and better institutional management. This fairly instrumentalist approach treats data literacy as a skillset with an emphasis on teachers learning techniques such as data handling.

On the other hand, as Raffaghelli and Stewart (this issue) argue, data literacy might be seen as a set of critical orientations which would allow educators and students to interrogate the claims accompanying data systems, question the validity or reliability of the data produced, and pose ethical challenges to the uptake and use

of data in education. Their extensive review of the available literature on educators' data literacies highlights significant gaps that stand to limit critical professional learning about datafication and its consequences. Loftus and Madden (this issue) extend a critical data literacies approach to the space of a computer science classroom. Specifically, they report on an experimental module on the Internet of Things where students can both explore the construction of machine learning models and learn to reflect on their social consequences as 'students who will be building the autonomous, connected systems of the future'. We might add here, too, that critical data literacy should be a requirement of university managers and administrators, as a counter to competitive, marketized forms of governing universities through their data trails.

Marketing and marketization

Papers across the special issue emphasize the ways that datafication is implicated in marketization. In a context where HE institutions are increasingly treated as competitive market actors, and subjected to continuous monitoring and comparison, datafication makes it ever more possible to quantify and analyse institutional performances at multiple scales (Komljenovic & Robertson, 2018). Datafication also enables institutions to generate data that demonstrates their competitive advantages over others, such as through data dashboards, league tables, and the presentation of reputation-raising data for marketing and branding.

Bamberger, Bronshtein and Yemini (this issue) make the astute observation that datafication has now become a key source of university marketing. The data feeds into particular portrayals of universities that are used to attract students, especially international students whose attention is solicited through online social media campaigns. Indeed, social media has become a key new space of university marketing and reputation management. To a significant extent, techniques of social media marketing have even fed back into universities own branding strategies, as institutions reconfigure their brands in order to circulate more effectively online. These marketing strategies are now an increasingly important aspect of HE marketization, as universities have become globally competitive actors seeking a share of the international student market.

But such data-led marketing may also exacerbate historical inequalities, as 'top-ranked' institutions that attract advantaged students maintain their 'bragging rights' and competitive position in league tables at the expense of institutions with missions not captured by the rankings. 'Because rankings are standardized algorithms applied to all schools, they punish schools that do not conform to the image of excellence embedded in the rankings' (Espeland & Sauder, 2016, 179), and therefore create new disadvantages and potential long-term inequalities for low-income or nonstandard students at such institutions.

Data inequalities

Datafication reinforces and reproduces historical inequalities. As Gilliard and Culik (2016) have compellingly shown, data systems systemically disadvantage certain social and minority groups through 'digital redlining', that is, the denial of services or the targeting of particular marginalized groups. Algorithmic decision-making automates inequalities, and discriminates along racialized and gendered lines. For example, controversy has arisen over automated recruitment systems, where applications for jobs are screened without human oversight, because they are found to disadvantage applicants from already under-represented groups, based on previous training data showing that predominantly white male applicants perform more highly.

Hayes (this issue) makes a positive move toward greater equity by arguing that 'epistemic equality' and non-discrimination should be officially considered as an indicator of teaching excellence in datafied and metricized systems of governance. Offering a practical example from the UK context, Hayes suggests that the often controversial systems designed to measure teaching excellence such as TEF, AHELO, CALOREE and CLA can be used to actively 're-shape the existing status quo with respect to who gets to produce knowledge' by being re-worked according to the principles of critical pedagogy. Such commitments to equity and non-discrimination, however, fit uneasily into systems of student monitoring that in some cases appear increasingly intrusive and surveillant.

Smart surveillance

The datafication of our private lives has normalised a 'liquid surveillance' (Bauman and Lyon 2013), in which the watching of ourselves and each other facilitated by digital technologies generates continuous flows of data about individuals, increasingly including young people (Lupton and Williamson 2017). This has been amplified and monetised by social media and other corporations built according to platform models which depend on the extraction, profiling and commercialisation of large amounts of user data to generate profit. Such 'surveillance capitalism' might be seen as one defining characteristic of our current technological moment (Zuboff 2019).

With sensor and device-based tracking of individuals technically possible, location analytics have potential – should universities wish it – for data-enabled student tracking, attendance and 'engagement' monitoring albeit with legal restrictions provided by data protection regulation. Quantification in education promises in the near future to extend to neurotechnological ways of understanding learning, with commercial educational technology initiatives promising new brain–computer interfaces, cognitive training tools and electronic neurostimulators. Facial recognition and emotion recognition technologies as means of mapping, tracking

and recording student engagement are already available and used by a small but growing number of universities, despite extensive dissent among academics and students. Fundamental questions emerge from this concerning the ownership of data, its ethical uses, permanence, the risks of reproducing discrimination, and implications for privacy and liberty of students and academics. It also raises questions about trust in students, as students are increasingly viewed as 'transmitters' of data than can be sensed from autonomic signals emitted from the body, rather than as sense-making actors who might be engaged in dialogue.

With increasing urbanisation, the datafication of cities and campuses is accelerating these new forms of 'data exhaust' offering potential for building new kinds of data-rich education, alongside risks for pulling universities into new forms of normalized, politicised, data-driven surveillance and monitoring. Kwet (this issue) interrogates the 'smart campus' as a 'new frontier' for the university, foregrounding the risks universities run by normalising surveillance architectures on-campus, calling for a ban on various forms of dataveillance and arguing for decentralised services, 'public interest technology' and more democratic pedagogic models which equip academic communities to push back on creeping surveillance and its modes of technocratic control. Furthermore, in such a data-intensive smart university, the possibilities for educational research are being reconfigured and, in some cases, displaced to new technical experts such as education data scientists and learning engineers who can straddle the social and computer sciences.

Social/data science

The final theme to raise here concerns methodology. As data has become the basis of increasingly influential practices of knowledge production, and data science has become both an academic discipline and a high-status professional occupation, a new debate has emerged about the status of the social sciences, arts and humanities. Sociology in particular has begun to reflect on its authority in comparison with the massive analyses of social data possible for commercial firms, leading to new proposals for 'digital sociology' to bridge the data science/social science divide (Marres, 2017). A digital sociology would not only mobilize big data for sociological analysis, but also critique datafication and its epistemological and ethical limits.

Selwyn and Gasevic (this issue) present a dialogue between educational data science and critical social science in an attempt to find a collaborative and productive interdisciplinary methodology for meaningful analysis of educational data. Doing so is likely to be an extremely long-term and arduous undertaking, as it requires navigating very different and highly specialized research problems, epistemologies and ontologies. However, critical social scientific researchers of educational data systems are already beginning to ask what might be learnt from a

closer productive engagement with educational data. Gorur (2020), for example, suggests that the relational approach of big data analyses might open up new opportunities for social analysis, particularly in actor-network theory approaches that focus on tracing complex associations between heterogeneous actors and objects. New forms of social/data science are, then an opportunity for disciplinary experimentation and creativity, but need to be enacted with reflexive attention to ongoing controversies over datafication.

Beyond critique?

The turn to data in higher education raises myriad problems for analysis, a selection of which are represented in the papers in this special issue. Although datafication itself has not determined the issues confronted in the papers—such as marketization, inequality, performance ranking, and so on—it is certainly reinforcing, reproducing and amplifying them. Data power is incorporated into the infrastructures, software, measures, and algorithms that constitute the digital architecture of HE.

As we grapple with these critical issues we should also be mindful of positive future opportunities both for research and for the remaking of universities as public institutions. Academics and students are beginning to address the challenge of how to resist these trends, both intellectually and practically, starting with a restatement of the inherent social and public good of higher education (The Analogue University, 2019). Connell (2019) has recently outlined a vision for a 'good university' in which the forces of corporate culture, academic capitalism and performative managerialism are rejected in favour of democratic, engaged, creative, and sustainable practices. Such efforts demonstrate the necessity both of engaged critical analysis and of moving beyond critique to practical engagements and mobilization against the datafied university. The datafication of teaching in higher education remains a key focus for future studies, and we hope as editors that the collection helps catalyse further conceptual and empirical interrogation of these fast-moving developments, as well as animating efforts by educators to reimagine and reshape the role of the university in the 2020s. Is it possible for the datafication of teaching to support the democratic mission of a good university?

References

ACE (American Council on Education). (2020). ACE Initiative will explore the potential for blockchain to close equity gaps across the education landscape. ACE, 6 February:

Ambrose, M. (2015) Lessons from the avalanche of numbers: big data in historical perspective. *I/S: A Journal of Law and Policy for the Information Society,* 11(2), 201-277.

Andrejevic, M. & Selwyn, N. (2019). Facial recognition technology in schools: critical questions and concerns. Learning, Media and Technology: https://doi.org/10.1080/17439884.2020.1686014

Bacevic, J. (2019). With or without U? Assemblage theory and (de)territorialising the university. *Globalisation, Societies and Education*, *17*(1), 78–91.

Bauman, Z., & Lyon, D. (2013). *Liquid surveillance: A conversation*. Hoboken: John Wiley & Sons.

Beer, D. (2019) *The Data Gaze*. London: Sage.

Beer, D. (2016). *Metric Power*. London: Palgrave Macmillan

Berry, D. (2018). The Data-Intensive University. Stunlaw, 14 September: http://stunlaw.blogspot.com/2018/09/the-data-intensive-university.html

Bigo, D., Isin, E. & Ruppert, E. (2019). Data Politics. In Bigo, D., Isin, E. & Ruppert, E. (eds). Data Politics: Worlds, subjects, rights (pp. 1-18). London: Routledge.

Bowker, G.C. (2008). *Memory Practices in the Sciences*. London: MIT Press.

Buckingham Shum, S. & Luckin, R. (2019). Learning analytics and AI: Politics, pedagogy and practices. British Journal of Educational Technology, 50(6), 2785-2793.

Burrows, R. (2012). Living with the h-index: Metric assemblages in the contemporary academy. *The Sociological Review,* 60(2), 355-372.

Busch, L. (2017). *Knowledge for Sale: The neoliberal takeover of higher education*. London: MIT Press.

Castaneda, L. & Selwyn, N. (2018). More than tools? Making sense of the ongoing digitizations of higher education. *International Journal of Educational Technology in Higher Education*, 15(22), https://doi.org/10.1186/s41239-018-0109-y

Cheney-Lippold, J. (2011) A new algorithmic identity: soft biopolitics and the modulation of control, *Theory, Culture & Society,* 28(6), 164-81.

Connell, R. (2019). *The Good University: What universities actually do and why it's time for radical change.* Melbourne: Monash University Press.

Couldry, N. & Meijas, U.A. (2019). Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television and New Media* 20(4), 336-349.

Crawford, K. & Paglen, T. (2019). Excavating AI: The Politics of Images in Machine Learning Training Sets. Excavating AI: https://www.excavating.ai/

Czerniewicz, L. (2018). Unbundling and Rebundling Higher Education in an Age of Inequality. *Educause Review*, November-December, 10-24: https://er.educause.edu/articles/2018/10/unbundling-and-rebundling-higher-education-in-an-age-of-inequality

Elish, M.C. & boyd, d. (2018). Situating methods in the magic of Big Data and AI. *Communication Monographs* 85(1), 57-80.

Espeland, W.N. & Sauder, M. (2016). *Engines of Anxiety: Academic rankings, reputation, and accountability*. New York: Russell Sage Foundation.

Esposito, E. & Stark, D. (2019). What's Observed in a Rating? Rankings as Orientation in the Face of Uncertainty. *Theory, Culture and Society*: https://doi.org/10.1177/0263276419826276

Fazackerley, A. (2020). Alarm at Ofsted-style plan to rank universities by graduate earnings. *The Guardian*, 11 February: https://www.theguardian.com/education/2020/feb/11/alarm-at-ofsted-style-plan-to-rank-universities-by-graduate-earnings

Foucault, M. (2007). *Security, Territory and Population: Lectures at the College de France 1977-1978*. Trans. G. Burchell. New York: Palgrave Macmillan.

Fourcade, M. & Healy, K. (2017). Seeing like a market. *Socio-economic Review*, 15(1), 9-29.

Gent, E. (2020). Is education the new currency? BBC Worklife, 11 February: https://www.bbc.com/worklife/article/20200210-is-education-the-new-currency

Gillard, C. & Culik, H. (2016). Digital Redlining, Access, and Privacy. Common Sense Education, 24 May: https://www.commonsense.org/education/articles/digital-redlining-access-and-privacy

Gorur, R. (2020). Afterword: embracing numbers? *International Studies in Sociology of Education*: https://doi.org/10.1080/09620214.2020.1720518

Jasanoff, S. (2018). Virtual, visible, and actionable: Data assemblages and the sightlines of justice. *Big Data and Society*: https://doi.org/10.1177/2053951717724477

Kitchin, R. (2014). *The Data Revolution: Big data, open data, data infrastructures and their consequences*. London: Sage.

Knox, J., Williamson, B. & Bayne, S. (2020). Machine behaviourism: future visions of 'learnification' and 'datafication' across humans and digital technologies. *Learning, Media and Technology*: https://doi.org/10.1080/17439884.2019.1623251

Komljenovic, J. & Robertson, S. (2016). The dynamics of 'market-making' in higher education. *Journal of Education Policy*, 31(5), 622–636.

Lane, J.E. (ed.) (2014). *Building a smarter university: Big data, innovation and analytics*. Albany: State University of New York Press

Lawn, M. (2013). The rise of data in education. In M. Lawn (ed.) *The Rise of Data in Education Systems: Collection, visualization and use*, 7-10. Oxford: Symposium.

Lupton D. & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society*, 19(5), 780-794.

Mackenzie, A. (2018). *Machine Learners: Archaeology of a data practice*. London: MIT Press.

Mandinach, E.B. & Gummer, E.S. (2016). Every teacher should succeed with data literacy. *Phi Delta Kappan* 97(8), 43-4.

Marres, N. (2017). *Digital Sociology: The reinvention of social research*. Cambridge: Polity.

McCowan, T. (2017). Higher education, unbundling, and the end of the university as we know it. *Oxford Review of Education,* 43(6), 733-748.

McStay, A. (2018). *Emotional AI: The rise of empathic media*. London: Sage.

Perrotta, C. & Selwyn, N. (2019). Deep learning goes to school: toward a relational understanding of AI in education. *Learning, Media and Technology*: https://doi.org/10.1080/17439884.2020.1686017

Piattoeva, N. & Boden, R. (2020). Escaping numbers? The ambiguities of the governance of education through data. *International Studies in Sociology of Education*: https://doi.org/10.1080/09620214.2020.1725590

Raley, R. (2013). Dataveillance and counterveillance. In L. Gitelman (ed.) *'Raw Data' is an Oxymoron*: 121-146. London: MIT Press.

Rose, N. (1999). *Powers of Freedom: Reframing political thought*. Cambridge: Cambridge University Press.

Ruppert, E. (2018). *Sociotechnical Imaginaries of Different Data Futures: An Experiment in Citizen Data.* Rotterdam, NL: Erasmus University Rotterdam.

Sellar, S. & Hogan, A. (2019). *Pearson 2025: Transforming teaching and privatising education data.* Brussels: Education International.

Smyth, J. (2017). *The Toxic University: Zombie Leadership, Academic Rock Stars and Neoliberal Ideology*. London: Palgrave Macmillan.

The Analogue University. (2019). Correlation in the Data University: Understanding and Challenging Targets-based Performance-management in Higher Education. *ACME: An International Journal for Critical Geographies*, 18(6), 1184-1206

Williamson, B. (2017). *Big Data in Education: The digital future of learning, policy and practice*. London: Sage.

Williamson, B. (2018). The hidden architecture of higher education: building a big data infrastructure for the 'smarter university'. *International Journal of Educational Technology in Higher Education*, 15(12): https://doi.org/10.1186/s41239-018-0094-1

Williamson, B. (2019). Policy networks, performance metrics and platform markets: Charting the expanding data infrastructure of higher education. *British Journal of Educational Technology*, 50(6), 2794-2809.

Williamson, B. (2020). Making markets through digital platforms: Pearson, edu-business, and the (e)valuation of higher education. *Critical Studies in Education*: https://doi.org/10.1080/17508487.2020.1737556

Wilsdon, J., et al. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. DOI: 10.13140/RG.2.1.4929.1363

Zeide, E. (2019). Robot Teaching, Pedagogy, and Policy. SSRN: https://ssrn.com/abstract=3441300

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The fight for a human future at the new frontier of power*. London: Profile Books.